< 과제 #2 : Numpy/Pandas/Matplotlib >

[A] NumPy

1) 모양이 (20,)인 4개의 1차원 배열 a, b, c, d를 만드시오.
   - ✓ a : 처음 15개는 3, 나머지 5개는 4인 배열
   - ✓ b : 정수 집합 {1, 2}에서 랜덤하게 뽑은 난수 배열(단, seed는 1로 설정)
   - ✓ c : 정수 집합 {1, 2, ...,, 99, 100}에서 랜덤하게 뽑은 난수 배열(단, seed는 2로 설정)
   - ✓ d : [0, 1) 구간에서 뽑은 난수 배열(단, seed는 3으로 설정)에 대하여 100을 곱한 후 소수점 이하는 버리고 정수 부분만 취하여 1을 더한 배열
2) 1)에서 만든 4개의 1차원 배열을 각각 2차원 배열, 즉 모양이 (20,1)이 되도록 재구조화한 후, 칼럼으로 이어 붙여 모양이 (20,4)인 2차원 배열 m을 만드시오.

```
array([[ 3.,  2., 41., 56.],
       [ 3.,  2., 16., 71.],
       [ 3.,  1., 73., 30.],
       [ 3.,  1., 23., 52.],
       [ 3.,  2., 44., 90.],
       [ 3.,  2., 83., 90.],
       [ 3.,  2., 76., 13.],
       [ 3.,  2.,  8., 21.],
       [ 3.,  2., 35.,  6.],
       [ 3.,  1., 50., 45.],
       [ 3.,  1., 96.,  3.],
       [ 3.,  2., 76., 46.],
       [ 3.,  1., 86., 65.],
       [ 3.,  2., 48., 28.],
       [ 3.,  2., 64., 68.],
       [ 4.,  1., 32., 60.],
       [ 4.,  1., 91.,  3.],
       [ 4.,  2., 21., 56.],
       [ 4.,  1., 38., 26.],
       [ 4.,  1., 40., 42.]])
```

3) 배열 m에 대하여 열별 평균을 구하시오.
4) 0번째 칼럼의 값이 3인, 즉 처음 15개의 행을 취하여 m3를 만들고 각 열의 평균을 구하시오.
5) 1번째 칼럼의 값이 1인 행을 취하여 m1을 만들고 각 열의 최댓값을 구하시오.
6) 3번째 열의 값이 2번째 열의 값보다 더 큰 행의 행 인덱스를 구하시오.

[B] Pandas

1) 타이타닉 데이터 파일을 읽어 titanic 데이터프레임을 만든 후, Name, Ticket, Cabin 칼럼을 삭제한 titanic2 데이터프레임을 만들어 마지막 3개의 행을 출력하시오.
2) titanic2에는 Age와 Embarked 칼럼에 결측치가 존재한다. Age의 결측치는 평균으로 대체하고, Embarked는 가장 많이 출현하는 값으로 대체하여 titanic3 데이터프레임을 만드시오. (참고 : fillna 메서드 도움말 예제)

```
>>> df = pd.DataFrame([[np.nan, 2, np.nan, 0],
...                    [3, 4, np.nan, 1],
...                    [np.nan, np.nan, np.nan, 5],
...                    [np.nan, 3, np.nan, 4]],
...                   columns=list("ABCD"))
>>> df
     A    B   C  D
0  NaN  2.0 NaN  0
1  3.0  4.0 NaN  1
2  NaN  NaN NaN  5
3  NaN  3.0 NaN  4
```

```
>>> values = {"A": 0, "B": 1, "C": 2, "D": 3}
>>> df.fillna(value=values)
     A   B   C   D
0   0.0 2.0 2.0 0
1   3.0 4.0 2.0 1
2   0.0 1.0 2.0 5
3   0.0 3.0 2.0 4
```

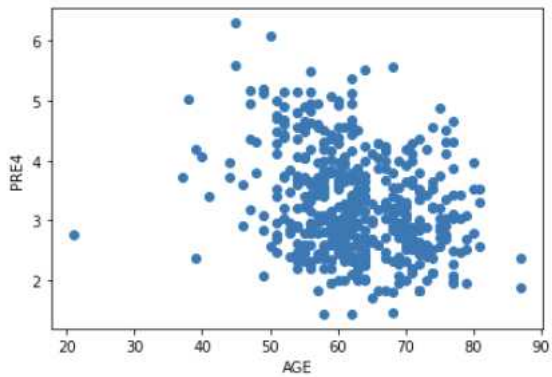
3) titanic3 데이터프레임에 대하여 아래 집계를 실시하시오.
   - ✓ Survived 그룹별 분포(개수)
   - ✓ Survived 그룹별 Age 평균
   - ✓ Pclass 그룹별 분포, 단 Pclass 번호순으로 정렬
   - ✓ Survived × Pclass 별 분포(개수)

4) titanic3으로부터 Age>70인 행만 추출하여 titanic4를 만들고, titanic으로부터 PassengerId와 Name만을 추출한 titanic4를 만들어 두 데이터프레임을 PassengerId를 키로 하여 내부조인하시오.

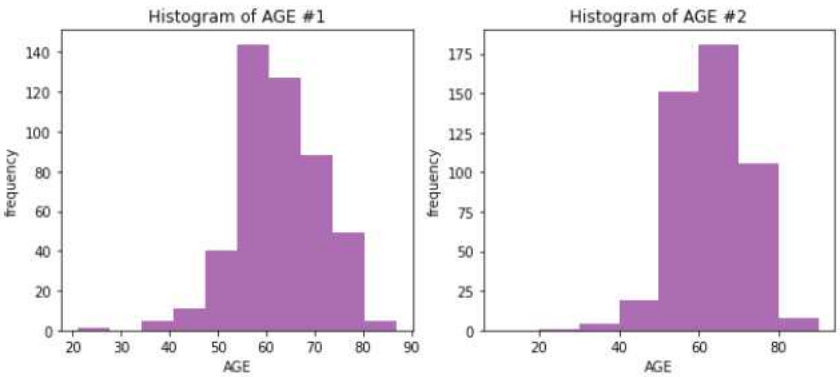
[C] Matplotlib

ThoracicSurgery.csv 파일은 폐암 수술 환자의 수술 전 진단 데이터(17개 : 종양의 유형, 폐활량, 호흡 곤란 여부, 고통 정도, 기침, 흡연, 천식 여부 등)와 수술 후 생존 결과(Risk1Yr)를 기록한 의료 기록 데이터이다. 파일을 읽어 데이터프레임을 생성한 후 아래 차트를 작성하시오.

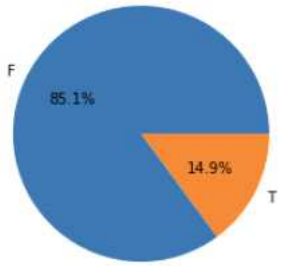
1) AGE와 PRE4의 산점도



2) AGE에 대한 히스토그램(#1:디폴트, #2:bins 설정(10대, 20대, ..., 90대))



3) Risk1Yr에 대한 파이차트(pie chart)



4) PRE14별 PRE5 평균의 막대그래프(bar chart)